A test of significance in functional quadratic regression [1]

Lajos Horváth and Ron Reeder

Department of Mathematics, University of Utah

Abstract

We consider a quadratic functional regression model in which a scalar response depends on a functional predictor; the common functional linear model is a special case. We wish to test the significance of the nonlinear term in the model. We develop a testing method which is based on projecting the observations onto a suitably chosen finite dimensional space using functional principal component analysis. The asymptotic behavior of our testing procedure is established. A simulation study shows that the testing procedure has good size and power with finite sample sizes. We then apply our test to a data set provided by Tecator, which consists of near-infrared absorbance spectra and fat content of meat.

Keywords: Absorption spectra; Asymptotics; Functional data analysis; Polynomial regression; Prediction; Principal component analysis.

1 Introduction and results

In a predictive model, it may be more natural and appropriate for certain quantities to be represented as trajectories rather than a single number (Kirkpatrick and Heckman, 1989). For example, a young animal's size may be considered as a function of time, giving a growth trajectory. A model to predict a certain response from growth trajectories is useful to animal breeders because they may be able to produce more valuable animals by changing their growth patterns (Fitzhugh, 1976). Müller and Zhang (2005) used egg-laying trajectories from Mediterranean fruit flies to predict a female fly's remaining lifetime. Frank and Friedman (1993) and Wold (1993) provide an early discussion on the applications of principal components to analyze curves in chemistry. Yao and Müller (2010) and Borggaard and Thodberg (1992) used absorbance trajectories to predict the fat content of meat samples. The absorbance at any particular wavelength is a measurement related to the proportion of light that passes through a meat sample. A representative sample of 15 of the 240 absorbance trajectories are pictured in Figure 1.

In functional regression, special attention has been given to functional linear models (Cardot et al., 2003; Shen and Faraway, 2004; Cai and Hall, 2006; Hall and Horowitz, 2007). However, it is pointed out in Yao and Müller (2010) that this model imposes a constraint on the regression relationship that may not be appropriate in some scenarios. Yao and Müller (2010) generalized this to a functional polynomial model, which has greater flexibility. In functional polynomial regression, as in standard polynomial regression, one must balance the costs and benefits of using more parameters in the model. In this paper, we will develop a

[1] Research partially supported by NSF grant DMS 0905400

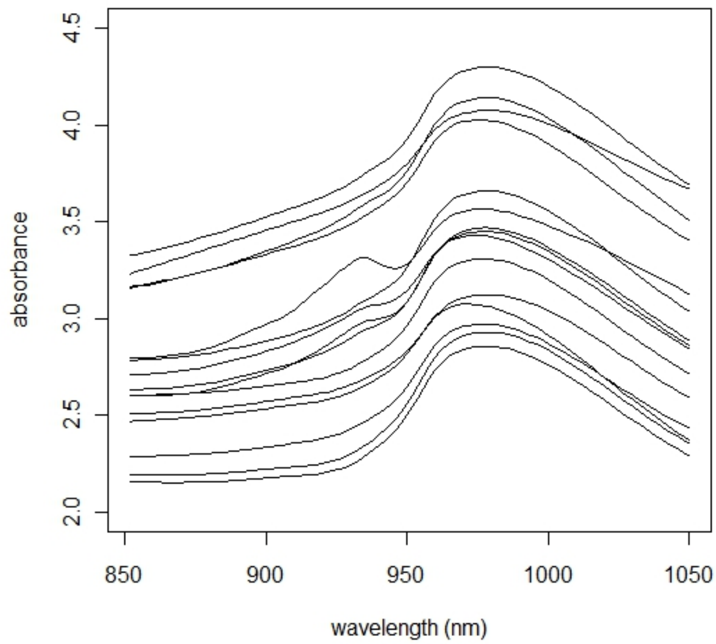

Figure 1: Absorbance trajectories from 15 samples of finely chopped pure meat.

test to determine if a quadratic term is justified in the model or if a functional linear model adequately describes the regression relationship.

The functional quadratic model in which a scalar response, Y_n, is paired with a functional predictor, $X_n(t)$, is defined as

$$Y_n = \mu + \int_0^1 k(t) X_n^c(t)\, dt + \int_0^1 \int_0^1 h(s,t) X_n^c(s) X_n^c(t)\, dt\, ds + \varepsilon_n, \tag{1.1}$$

where $X_n^c(t) = X_n(t) - E\left(X_n(t)\right)$ is the centered predictor process. If $h(s,t) = 0$, then $\mu = E(Y_n)$ and (1.1) reduces to the functional linear model

$$Y_n = \mu + \int_0^1 k(t) X_n^c(t)\, dt + \varepsilon_n. \tag{1.2}$$

Cardot and Sarda (2011) and Mas and Pumo (2011) point out in their survey papers that since we can choose a function in (1.2), the functional linear model can be used in a large variety of applications. The functional linear model provides a very simple relation between $X_n(t)$ and Y_n, so it is important to check if the more involved quadratic model (1.1) provides a real improvement. In other words, one should test whether the quadratic term is really needed. To test the significance of the quadratic term in (1.1), we test the null hypothesis,

$$H_0 : h(s,t) = 0, \tag{1.3}$$

against the alternative

$$H_A : h(s,t) \neq 0.$$

To reduce the dimensionality and avoid overfitting in our functional regression model, we will project the predictor process onto a suitably chosen finite dimensional space. The space is spanned by the eigenfunctions of $C(t,s) = E(X_n(t) - \mu_X(t))(X_n(s) - \mu_X(s))$, the covariance function of the predictor process, where $\mu_X(t) = EX_n(t)$. We will denote the eigenfunctions and associated eigenvalues by $\{(v_i(t), \lambda_i),\ 1 \le i \le \infty\}$. We can and will assume that λ_i is the i^{th} largest eigenvalue and that the eigenfunctions are orthonormal. It is clear that we can assume that h is symmetric, and we also impose the condition that the kernels are in L^2:

$$h(s,t) = h(t,s) \ \text{ and} \int_0^1 \int_0^1 h^2(s,t)dtds < \infty, \tag{1.4}$$

$$\int_0^1 k^2(t)dt < \infty. \tag{1.5}$$

Thus we have the expansions

$$\begin{aligned} h(s,t) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{i,j}v_j(s)v_i(t) \\ &= \sum_{i=1}^{\infty} a_{i,i}v_i(s)v_i(t) + \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} a_{i,j}\left(v_j(s)v_i(t) + v_i(s)v_j(t)\right) \end{aligned} \tag{1.6}$$

and

$$k(t) = \sum_{i=1}^{\infty} b_i v_i(t). \tag{1.7}$$

By projecting onto the space spanned by $\{v_1, \dots, v_p\}$ and using (1.6) and (1.7), we can write the model (1.1) as

$$Y_n = \mu + \sum_{i=1}^{p} b_i \langle X_n^c, v_i\rangle + \sum_{i=1}^{p}\sum_{j=i}^{p}(2 - 1\{i=j\})a_{i,j}\langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle + \varepsilon_n^*, \tag{1.8}$$

where

$$\varepsilon_n^* = \varepsilon_n + \sum_{i=p+1}^{\infty} b_i\langle X_n^c, v_i\rangle + \sum_{i=p+1}^{\infty}\sum_{j=i}^{\infty}(2-1\{i=j\})a_{i,j}\langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle + \sum_{i=1}^{p}\sum_{j=p+1}^{\infty} 2a_{i,j}\langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle.$$

We note that (1.8) is written as a standard linear model, but the error term, ε_n^*, and the design points, $\{\langle X_n^c, v_i\rangle, 1 \le i \le p\}$, are dependent.

Unfortunately, we cannot use (1.8) directly for statistical inference since $v_i(t)$ and $\mu_X(t)$ are unknown. We estimate $\mu_X(t)$ and $C(t,s)$ with the corresponding empiricals

$$\bar{X}(t) = \frac{1}{N}\sum_{n=1}^{N} X_n(t)$$

and

$$\hat{C}(t,s) = \frac{1}{N}\sum_{n=1}^{N}\left(X_n(t) - \bar{X}(t)\right)\left(X_n(s) - \bar{X}(s)\right).$$

The eigenvalues and the corresponding eigenfunctions of $\hat{C}(t,s)$ are denoted by $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots$ and $\hat{v}_1, \hat{v}_2, \dots$. Eigenfunctions corresponding to unique eigenvalues are uniquely determined up to signs. For this reason, we cannot expect more than to have $\hat{c}_i \hat{v}_i(t)$ be close to $v_i(t)$, where the $\hat{c}_i$'s are random signs. We replace equation (1.8) with

$$Y_n = \mu + \sum_{i=1}^{p} b_i \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle + \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1\{i=j\}) a_{i,j} \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{c}_j \hat{v}_j \rangle + \varepsilon_n^{**}, \quad (1.9)$$

where

$$\begin{aligned} \varepsilon_n^{**} =& \varepsilon_n^* + \sum_{i=1}^{p} b_i \langle X_n^c, v_i - \hat{c}_i \hat{v}_i \rangle + \sum_{i=1}^{p} b_i \langle \bar{X} - \mu_X, \hat{c}_i \hat{v}_i \rangle \\ & - \sum_{i=1}^{p} \sum_{j=i}^{p} (2 - 1\{i=j\}) a_{i,j} \left(\langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{c}_j \hat{v}_j \rangle - \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle \right). \end{aligned}$$

We can write (1.9) in the concise form

$$\mathbf{Y} = \hat{\mathbf{Z}} \begin{pmatrix} \tilde{\mathbf{A}} \\ \tilde{\mathbf{B}} \\ \mu \end{pmatrix} + \boldsymbol{\varepsilon}^{**}, \quad (1.10)$$

where

$$\begin{aligned} \mathbf{Y} &= \left(Y_1, Y_2, \dots, Y_N \right)^T, & \tilde{\mathbf{A}} &= \operatorname{vech}\left(\{ \hat{c}_i \hat{c}_j a_{i,j} \left(2 - 1\{i=j\} \right), \ 1 \leq i \leq j \leq p \}^T \right), \\ \tilde{\mathbf{B}} &= \left(\hat{c}_1 b_1, \hat{c}_2 b_2, \dots, \hat{c}_p b_p \right)^T, & \boldsymbol{\varepsilon}^{**} &= \left(\varepsilon_1^{**}, \varepsilon_2^{**}, \dots, \varepsilon_N^{**} \right)^T, \end{aligned}$$

and

$$\hat{\mathbf{Z}} = \begin{pmatrix} \hat{\mathbf{D}}_1^T & \hat{\mathbf{F}}_1^T & 1 \\ \hat{\mathbf{D}}_2^T & \hat{\mathbf{F}}_2^T & 1 \\ \vdots & \vdots & \vdots \\ \hat{\mathbf{D}}_N^T & \hat{\mathbf{F}}_N^T & 1 \end{pmatrix}$$

with

$$\begin{aligned} \hat{\mathbf{D}}_n &= \operatorname{vech}\left(\{ \langle \hat{v}_i, X_n - \bar{X} \rangle \langle \hat{v}_j, X_n - \bar{X} \rangle, \ 1 \leq i \leq j \leq p \}^T \right), \\ \hat{\mathbf{F}}_n &= \left(\langle X_n - \bar{X}, \hat{v}_1 \rangle, \quad \langle X_n - \bar{X}, \hat{v}_2 \rangle, \quad \dots, \quad \langle X_n - \bar{X}, \hat{v}_p \rangle \right)^T. \end{aligned}$$

The half-vectorization, vech$(\cdot)$, stacks the columns of the lower triangular portion of the matrix under each other. Although we write our model in the form of a general linear model, it is important to note that it is not a classical linear model. First, $\boldsymbol{\varepsilon}^{**}$ is correlated with $\hat{\mathbf{Z}}$ because $\boldsymbol{\varepsilon}^{**}$ contains additional error terms which come from projecting onto a p-dimensional space. Another important difference between (1.10) and a classical linear model is that the parameters to be estimated, $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$, are random; they depend on the random signs, $\hat{c}_i$. We estimate $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and μ using the least squares estimator:

$$\begin{pmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} \\ \hat{\mu} \end{pmatrix} = \left(\hat{\mathbf{Z}}^T \hat{\mathbf{Z}} \right)^{-1} \hat{\mathbf{Z}}^T \mathbf{Y}. \quad (1.11)$$

To represent elements of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, we will use the notation that $\hat{\mathbf{A}} = \mathrm{vech}(\{\hat{a}_{i,j}(2 - 1\{i = j\}), 1 \le i \le j \le p\}^T)$ and $\hat{\mathbf{B}} = \begin{pmatrix} \hat{b}_1, & \hat{b}_2, & \ldots, & \hat{b}_p \end{pmatrix}^T$.

We expect, under H_0, that $\hat{\mathbf{A}}$ will be close to zero since $\tilde{\mathbf{A}}$ is zero. If H_0 is not correct, we expect the magnitude of $\hat{\mathbf{A}}$ to be relatively large. This suggests that a testing procedure could be based on $\hat{\mathbf{A}}$. Due to the random signs coming from the estimation of the eigenfunctions, $\hat{\mathbf{A}}$ will not be asymptotically normal. However, if the random signs are "taken out," asymptotic normality can be established. Hence our test statistic will be a quadratic form of $\hat{\mathbf{A}}$ with some random weight matrices. Let

$$\hat{\mathbf{G}} = \frac{1}{N}\sum_{n=1}^{N} \hat{\mathbf{D}}_n \hat{\mathbf{D}}_n^T,$$

$$\hat{\mathbf{M}} = \frac{1}{N}\sum_{n=1}^{N} \hat{\mathbf{D}}_n,$$

and

$$\hat{\tau}^2 = \frac{1}{N}\sum_{n=1}^{N} \hat{\varepsilon}_n^2,$$

where

$$\hat{\varepsilon}_n = Y_n - \hat{\mu} - \sum_{i=1}^{p} \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle - \sum_{i=1}^{p}\sum_{j=i}^{p} (2 - 1\{i = j\}) \hat{a}_{i,j} \langle X_n - \bar{X}, \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{v}_j \rangle$$

are the residuals under H_0. We reject the null hypothesis if

$$U_N = \frac{N}{\hat{\tau}^2} \hat{\mathbf{A}}^T (\hat{\mathbf{G}} - \hat{\mathbf{M}}\hat{\mathbf{M}}^T) \hat{\mathbf{A}}$$

is large. The main result of this paper is the asymptotic distribution of U_N under the null hypothesis. First, we discuss the assumptions needed to establish asymptotics for U_N:

Assumption 1.1. $\{X_n(t), n \ge 1\}$ *is a sequence of independent, identically distributed Gaussian processes.*

Assumption 1.2.

$$E\left(\int_0^1 X_n^2(t)\, dt\right)^4 < \infty.$$

Assumption 1.3. $\{\varepsilon_n\}$ *is a sequence of independent, identically distributed random variables satisfying* $E\varepsilon_n = 0$ *and* $E\varepsilon_n^4 < \infty$*,*

and

Assumption 1.4. *the sequences* $\{\varepsilon_n\}$ *and* $\{X_n(t)\}$ *are independent.*

The last condition is standard in functional data analysis. It implies that the eigenfunctions $v_1, v_2, \ldots, v_p$ are unique up to a sign.

Assumption 1.5.

$$\lambda_1 > \lambda_2 > \cdots > \lambda_{p+1}.$$

Theorem 1.1. *If H_0, (1.5) and Assumptions 1.1–1.5 are satisfied, then*

$$U_N \xrightarrow{\mathcal{D}} \chi^2(r),$$

where $r = p(p+1)/2$ is the dimension of the vector $\hat{\mathbf{A}}$.

The proof of Theorem 1.1 is given in Section 4.

Remark 1.1. *By the Karhunen-Loève expansion, every centered, square integrable process, $X_n^c(t)$, can be written as*

$$X_n^c(t) = \sum_{\ell=1}^{\infty} \xi_{n,\ell}\varphi_\ell(t),$$

where φ_ℓ are orthonormal functions. Assumption 1.1 can be replaced with the requirement that $\xi_{n,1}$, $\xi_{n,2}$, ..., $\xi_{n,p}$ are independent with $E\xi_{n,\ell}^3 = 0$ and $E\xi_{n,\ell} = 0$ for all $1 \le \ell \le p$.

Our last result provides a simple condition for the consistency of the test based on U_N. Let $\mathbf{A} = \mathrm{vech}(\{a_{i,j}(2 - 1\{i = j\}),\ 1 \le i \le j \le p\}^T)$, i.e. the first $r = p(p+1)/2$ coefficients in the expansion of h in (1.6).

Theorem 1.2. *If (1.4), (1.5), Assumptions 1.1–1.5 are satisfied and $\mathbf{A} \ne \mathbf{0}$, then we have that*

$$U_N \xrightarrow{P} \infty.$$

The condition $\mathbf{A} \ne \mathbf{0}$ means that h is not the 0 function in the space spanned by the functions $v_i(t)v_j(s), 1 \le i, j \le p$.

2 A simulation study

In this section, we investigate the empirical size and power of the testing procedure for finite sample sizes. Seeking to obtain a test of size $\alpha = .01$, .05, or .10, a rejection region was chosen according to the limiting distribution of the test statistic. Since the limiting distribution is $\chi^2(r)$, the rejection region is (Δ, ∞), where $P\left(\chi^2(r) > \Delta\right) = \alpha$. Simulated data was then used to compute the outcome of the test statistic. Iterating this procedure 5,000 times, we kept track of the proportion of times that the outcome fell in the predetermined rejection region. When simulations are done under H_0, this gives us the empirical size of the test, which we expect to be close to the nominal size, α, for large sample sizes. When simulations are done under the alternative, H_A, the proportion gives us the empirical power of the test.

In our first simulation study, the ε_n's were generated according to the distribution of independent standard normals. We generated the $X_n(t)$'s according to the distribution of independent standard Brownian motions. Then, using $k(t) = 1$ and $h(s,t) = c$, we obtained Y_n according to (1.1). Thus the power of the test is a function of the parameter c. In particular, when $c = 0$, the null hypothesis is true. The resulting empirical size and power are given in Table 1.

The distribution of our test statistic has been shown to converge to a $\chi^2(r)$. Thus we expect the empirical and nominal size to be close for samples of size $N = 200$ and even closer when $N = 500$, as observed in Table 1. Since our testing procedure depends on the choice of how many principal components to keep, results are given in Table 1 for $p = 1$, 2, and 3. One possible method of selecting p is to follow the advice of Ramsay and Silverman (2005) and choose p so that approximately 85% of the variance within a sample is described by the first p principal components.

Although Theorem 1.1 is proven under the assumption that $X_n(t)$ is a Gaussian process, the result of Theorem 1.1 holds under relaxed conditions as discussed in Remark 1.1. We will now investigate the empirical size and power of our test when $X_n(t)$ is not a Gaussian process. We generate the ε_n's according to a uniform distribution on $(-0.5, 0.5)$. The predictors, $X_n(t)$, are generated according to $X_n(t) = \left(T_{1,n} + T_{2,n}t + T_{3,n}(2t^2 - 1) + T_{4,n}(4t^3 - 3t)\right)/4$, where $\{T_{i,n}, 1 \leq i \leq 4, 1 \leq n\}$ are iid random variables having a t-distribution with 5 degrees of freedom. The polynomials in the definition of $X_n(t)$ are the orthogonal Chebyshev polynomials. The resulting empirical size and power are given in Table 2. We see from Table 2 that our testing procedure is robust against non-Gaussian observations. Comparing Tables 1 and 2, we see that the value of the test statistics tends to be larger if the X_n's are not normally distributed for small N. The overrejection fades as N gets larger so in case of non-Gaussian X_n's, larger sample sizes are needed. This also explains the somewhat better power of the procedure in the case of non-Gaussian errors.

3 Application to spectral data

In this section we apply our test to the data set collected by Tecator and available at http://lib.stat.cmu.edu/datasets/tecator. Tecator used 240 samples of finely chopped pure meat with different fat contents. For each sample of meat, a 100 channel spectrum of absorbances was recorded using a Tecator Infratec food and feed analyzer. These absorbances can be thought of as a discrete approximation to the continuous record, $X_n(t)$. Also, for each sample of meat, the fat content, Y_n was measured by analytic chemistry.

The absorbance curve measured from the n^{th} meat sample is given by $X_n(t) = \log_{10}(I_0/I)$, where t is the wavelength of the light, I_0 is the intensity of the light before passing through the meat sample, and I is the intensity of the light after it passes through the meat sample. The Tecator Infratec food and feed analyzer measured absorbance at 100 different wavelengths between 850 and 1050 nanometers. This gives the values of $X_n(t)$ on a discrete grid from which we can use cubic splines to interpolate the values anywhere within the interval. A representative sample of 15 of the 240 absorbance trajectories are pictured in Figure 1.

Yao and Müller (2010) proposed using a functional quadratic model to predict the fat content, Y_n, of a meat sample based on its absorbance spectrum, $X_n(t)$. We are interested in determining whether the quadratic term in (1.1) is needed by testing its significance for this data set. From the data, we calculate U_{240}. The p-value is then $P\left(\chi^2(r) > U_{240}\right)$. The test statistic and hence the p-value are influenced by the number of principal components that we choose to keep. If we select p according to the advice of Ramsay and Silverman (2005), we will keep only $p = 1$ principal component because this explains more than 85% of the variation between absorbance curves in the sample. Table 3 gives p-values obtained using $p = 1$, 2,

and 3 principal components, which strongly supports that the quadratic regression provides a better model for the Tecator data.

Table 1: Empirical power of test (in %) based on 5,000 simulations using iid Brownian motions for $X_n(t)$ and iid standard normals for ε_n.

c	$\alpha = .01$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	1.02	1.37	1.95	1.10	1.30	1.15
0.2	10.81	6.87	6.52	30.35	20.35	12.85
0.4	49.51	37.24	29.76	91.90	84.25	74.35
0.6	86.68	77.74	70.19	100.00	99.70	98.75
0.8	98.50	96.05	92.98	100.00	100.00	100.00
1.0	99.94	99.57	99.05	100.00	100.00	100.00

c	$\alpha = .05$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	5.15	6.00	7.44	5.60	5.75	6.05
0.2	25.90	19.17	18.02	53.05	40.00	31.35
0.4	72.10	60.31	50.38	97.90	93.70	88.55
0.6	95.21	90.43	85.77	100.00	99.90	99.60
0.8	99.60	98.90	97.60	100.00	100.00	100.00
1.0	99.99	99.87	99.84	100.00	100.00	100.00

c	$\alpha = .10$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	10.27	11.18	13.35	10.60	11.05	11.55
0.2	36.60	29.50	27.03	65.00	52.45	43.75
0.4	80.89	71.08	62.27	99.30	96.60	93.10
0.6	97.60	94.77	90.91	100.00	99.95	99.75
0.8	99.85	99.47	98.57	100.00	100.00	100.00
1.0	99.99	99.95	99.91	100.00	100.00	100.00

Table 2: Empirical power of test (in %) based on 5,000 simulations using non-Gaussian $X_n(t)$ and non-normal ε_n.

c	$\alpha = .01$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	2.40	1.20	1.85	1.75	1.45	1.35
0.2	57.70	46.75	37.50	93.75	90.30	82.55
0.4	96.90	95.55	91.20	100.00	100.00	100.00
0.6	99.90	100.00	99.70	100.00	100.00	100.00
0.8	100.00	100.00	100.00	100.00	100.00	100.00
1.0	100.00	100.00	100.00	100.00	100.00	100.00

c	$\alpha = .05$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	8.00	5.75	8.15	7.20	5.45	6.10
0.2	74.50	64.55	56.45	98.55	96.30	92.00
0.4	99.40	98.35	96.55	100.00	100.00	100.00
0.6	99.95	100.00	99.85	100.00	100.00	100.00
0.8	100.00	100.00	100.00	100.00	100.00	100.00
1.0	100.00	100.00	100.00	100.00	100.00	100.00

c	$\alpha = .10$					
	$N = 200$			$N = 500$		
	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
0.0	13.60	12.15	14.60	13.60	10.35	11.50
0.2	82.30	74.25	65.55	98.90	97.70	95.25
0.4	99.65	99.10	97.95	100.00	100.00	100.00
0.6	99.95	100.00	99.90	100.00	100.00	100.00
0.8	100.00	100.00	100.00	100.00	100.00	100.00
1.0	100.00	100.00	100.00	100.00	100.00	100.00

Table 3: p-values (in %) obtained by applying our testing procedure to the Tecator data set with $p = 1$, 2, and 3 principal components.

p	1	2	3
p-value	1.25	13.15	0.00

4 Proof of Theorem 1.1

Proof of Theorem 1.1. We have from (1.10) and (1.11) that

$$\begin{aligned}\begin{pmatrix}\hat{\mathbf{A}}\\ \hat{\mathbf{B}}\\ \hat{\mu}\end{pmatrix} &= \left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\left(\hat{\mathbf{Z}}\begin{pmatrix}\tilde{\mathbf{A}}\\ \tilde{\mathbf{B}}\\ \mu\end{pmatrix} + \boldsymbol{\varepsilon}^{**}\right)\\ &= \begin{pmatrix}\tilde{\mathbf{A}}\\ \tilde{\mathbf{B}}\\ \mu\end{pmatrix} + \left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}.\end{aligned} \tag{4.1}$$

We also note that, under the null hypothesis, $a_{i,j} = 0$ for all i and j and therefore ε_n^* and ε_n^{**} of (1.8) and (1.9) reduce to

$$\varepsilon_n^* = \varepsilon_n + \sum_{i=p+1}^{\infty} b_i\langle X_n^c, v_i\rangle$$

and

$$\varepsilon_n^{**} = \varepsilon_n^* + \sum_{i=1}^{p} b_i\langle X_n^c, v_i - \hat{c}_i\hat{v}_i\rangle + \sum_{i=1}^{p} b_i\langle \bar{X} - \mu_X, \hat{c}_i\hat{v}_i\rangle.$$

To obtain the limiting distribution of $\sqrt{N}\hat{\mathbf{A}}$, we need to consider the vector $\sqrt{N}\left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$. We will show in Lemmas 6.2–6.7 that

$$\left(\left(\frac{\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}}{N}\right) - \begin{pmatrix}\boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & \mathbf{M}\\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda} & \mathbf{0}_{p\times 1}\\ \mathbf{M}^T & \mathbf{0}_{1\times p} & 1\end{pmatrix}\right) = o_P(1), \tag{4.2}$$

where $\boldsymbol{\zeta}$ is an unobservable matrix of random signs, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, $\mathbf{M} = E(\mathbf{D}_n)$, and $\mathbf{G} = E\left(\mathbf{D}_n\mathbf{D}_n^T\right)$, where

$$\mathbf{D}_n = \mathrm{vech}\left(\{\langle v_i, X_n^c\rangle\langle v_j, X_n^c\rangle,\ 1 \le i \le j \le p\}^T\right).$$

We see from (4.2) that the vector $\sqrt{N}\left(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$ has the same limiting distribution as

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\begin{pmatrix}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & -\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda}^{-1} & \mathbf{0}_{p\times 1}\\ -\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1} & \mathbf{0}_{1\times p} & 1+\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\mathbf{M}\end{pmatrix}\begin{pmatrix}\hat{\mathbf{D}}_n\\ \hat{\mathbf{F}}_n\\ 1\end{pmatrix}. \tag{4.3}$$

Since we are only interested in $\sqrt{N}\hat{\mathbf{A}}$ we need only consider the first $r = p(p+1)/2$ elements of the vector in (4.3). In Lemma 6.8 we show that these are given by

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} \quad \mathbf{0}_{r\times p} \quad -\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right)\begin{pmatrix}\hat{\mathbf{D}}_n\\ \hat{\mathbf{F}}_n\\ 1\end{pmatrix}\\
&\quad=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right)\\
&\quad=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right).
\end{aligned}$$

Then, in Lemma 6.9 we prove that

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right)\xrightarrow{\mathcal{D}} N\left(0,\tau^2\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\right),$$

where $\tau^2 = \mathrm{var}\left(\varepsilon_1^*\right)$. Finally, in Lemmas 6.10 and 6.11, we show that $\hat{\tau}^2 - \tau^2 = o_P(1)$. As a consequence of (4.2), we see that $\left(\hat{\mathbf{G}} - \hat{\mathbf{M}}\hat{\mathbf{M}}^T\right) - \boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)\boldsymbol{\zeta} = o_P(1)$. Since $\boldsymbol{\zeta}$ is a diagonal matrix of signs, $\boldsymbol{\zeta}\boldsymbol{\zeta} = I$, completing the proof of Theorem 1.1. □

5 Proof of Theorem 1.2

We provide only an outline of the proof since it follows the arguments used in the proof of Theorem 1.1. However, the arguments are simple since instead of obtaining an asymptotic limit distribution we only establish the weak law

$$\hat{\mathbf{A}}^T(\hat{\mathbf{G}}-\hat{\mathbf{M}}\hat{\mathbf{M}}^T)\hat{\mathbf{A}} \quad \xrightarrow{P} \quad \mathbf{A}^T(\mathbf{G}-\mathbf{M}\mathbf{M}^T)\mathbf{A}, \tag{5.1}$$

where $\mathbf{A} = \mathrm{vech}\left(\{a_{i,j}\left(2-1\{i=j\}\right),\ 1\le i\le j\le p\}^T\right)$ is like the vector $\tilde{\mathbf{A}}$ except without the random signs.

First we note that according to Lemma 6.1, the estimation of $v_1,\ldots,v_p$ by $\hat{v}_1,\ldots,\hat{v}_p$ causes only the introduction of the random signs $\hat{c}_1,\ldots,\hat{c}_p$. As in the proof of Theorem 1.1 one can verify that

$$\hat{\mathbf{A}} - \boldsymbol{\zeta}\mathbf{A} \quad \xrightarrow{P} \quad \mathbf{0}.$$

Lemmas 6.2 and 6.6 hold under H_0 as well as under H_A. This gives

$$\hat{\mathbf{G}} - \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} = o_P(1)$$

and

$$\hat{\mathbf{M}}\hat{\mathbf{M}}^T - \boldsymbol{\zeta}\mathbf{M}\mathbf{M}^T\boldsymbol{\zeta} = o_P(1),$$

completing the proof of (5.1).

6 Technical lemmas

Throughout the proofs in this section we will use $\|\cdot\|_1$ to be the 1-norm and $\|\cdot\|_2$ to be 2-norm on the unit interval, square, cube, or hypercube. The null hypothesis, H_0, is assumed throughout this section. We will make frequent use of the following lemma, which is established in Dauxois et al. (1982) and Bosq (2000).

Lemma 6.1. *If Assumptions 1.1, 1.2, and 1.5 hold, then*

$$\|\hat{c}_i\hat{v}_i(t) - v_i(t)\| = O_P\left(N^{-1/2}\right)$$

for each $1 \le i \le p$.

Lemma 6.2. *If Assumptions 1.1, 1.2, and 1.5 hold, then there is a non-random matrix* $\mathbf{G}$ *such that*

$$\left(\hat{\mathbf{G}} - \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta}\right) = o_P(1),$$

where $\hat{\mathbf{G}} = N^{-1}\sum_{n=1}^N \hat{\mathbf{D}}_n\hat{\mathbf{D}}_n^T$ *and* $\boldsymbol{\zeta} = \operatorname{diag}\left(\operatorname{vech}(\{\hat{c}_i\hat{c}_j, 1 \le i \le j \le p\}^T)\right)$.

Proof. By the Karhunen-Loéve expansion we have

$$X_n^c(t) = \sum_{\ell=1}^{\infty} \lambda_\ell^{1/2}\xi_\ell^{(n)}v_\ell(t). \tag{6.1}$$

Therefore an element of $\mathbf{D}_n\mathbf{D}_n^T$ is of the form $\sqrt{\lambda_i\lambda_j\lambda_k\lambda_\ell}\xi_i^{(n)}\xi_j^{(n)}\xi_k^{(n)}\xi_\ell^{(n)}$. Hence using the strong law of large numbers we conclude

$$\frac{1}{N}\sum_{n=1}^N \mathbf{D}_n\mathbf{D}_n^T \xrightarrow{a.s.} \mathbf{G},$$

where $\mathbf{G} = E\left(\mathbf{D}_n\mathbf{D}_n^T\right)$. Thus it suffices to show that

$$\frac{1}{N}\sum_{n=1}^N \left(\boldsymbol{\zeta}\hat{\mathbf{D}}_n\hat{\mathbf{D}}_n^T\boldsymbol{\zeta} - \mathbf{D}_n\mathbf{D}_n^T\right) = o_P(1). \tag{6.2}$$

Expressing (6.2) elementwise, we obtain

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^N \Big(&\langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\langle X_n - \bar{X}, \hat{c}_j\hat{v}_j\rangle\langle X_n - \bar{X}, \hat{c}_k\hat{v}_k\rangle\langle X_n - \bar{X}, \hat{c}_\ell\hat{v}_\ell\rangle \\ &- \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\langle X_n^c, v_k\rangle\langle X_n^c, v_\ell\rangle\Big) = o_P(1).\end{aligned} \tag{6.3}$$

In order to prove (6.3), it is enough to show that

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^N \Big(&\langle X_n^c, \hat{c}_i\hat{v}_i\rangle\langle X_n^c, \hat{c}_j\hat{v}_j\rangle\langle X_n^c, \hat{c}_k\hat{v}_k\rangle\langle X_n^c, \hat{c}_\ell\hat{v}_\ell\rangle \\ &- \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\langle X_n^c, v_k\rangle\langle X_n^c, v_\ell\rangle\Big) = o_P(1)\end{aligned} \tag{6.4}$$

and

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X},\hat{c}_j\hat{v}_j\rangle\langle X_n-\bar{X},\hat{c}_k\hat{v}_k\rangle\langle X_n-\bar{X},\hat{c}_\ell\hat{v}_\ell\rangle \\ -\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\langle X_n^c,\hat{c}_j\hat{v}_j\rangle\langle X_n^c,\hat{c}_k\hat{v}_k\rangle\langle X_n^c,\hat{c}_\ell\hat{v}_\ell\rangle\Big)=o_P\left(1\right). \tag{6.5}$$

We only establish (6.4), since the proof of (6.5) is essentially the same. Using Hölder's inequality, we obtain

$$\begin{aligned}
&\Bigg|\int_0^1\int_0^1\int_0^1\int_0^1\left(\frac{1}{N}\sum_{n=1}^{N}X_n^c(s)X_n^c(t)X_n^c(u)X_n^c(w)\right)\\
&\qquad\times\left(\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right)\,ds\,dt\,du\,dw\Bigg|\\
&\leq\left|\left|\frac{1}{N}\sum_{n=1}^{N}X_n^c(s)X_n^c(t)X_n^c(u)X_n^c(w)\right|\right|_2\\
&\qquad\times\left|\left|\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right|\right|_2.
\end{aligned}$$

By the law of large numbers in Hilbert spaces (cf. (Bosq, 2000)), we have that

$$\left|\left|\frac{1}{N}\sum_{n=1}^{N}X_n^c(s)X_n^c(t)X_n^c(u)X_n^c(w)\right|\right|_2=O_P\left(1\right),$$

so it remains only to show that

$$\left|\left|\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right|\right|_2=o_P\left(1\right).$$

Using Minkowski's inequality, Fubini's Theorem, the fact that $\|\hat{v}_i\|_2=\|v_i\|_2=1$, and then Lemma 6.1, we obtain

$$\begin{aligned}
&\left|\left|\hat{c}_i\hat{v}_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)-v_i(s)v_j(t)v_k(u)v_\ell(w)\right|\right|_2\\
&\quad\leq\left|\left|\left(\hat{c}_i\hat{v}_i(s)-v_i(s)\right)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\hat{c}_\ell\hat{v}_\ell(w)\right|\right|_2\\
&\qquad+\left|\left|v_i(s)\hat{c}_j\hat{v}_j(t)\hat{c}_k\hat{v}_k(u)\left(\hat{c}_\ell\hat{v}_\ell(w)-v_\ell(w)\right)\right|\right|_2\\
&\qquad+\left|\left|v_i(s)\hat{c}_j\hat{v}_j(t)\left(\hat{c}_k\hat{v}_k(u)-v_k(u)\right)v_\ell(w)\right|\right|_2\\
&\qquad+\left|\left|v_i(s)\left(\hat{c}_j\hat{v}_j(t)-v_j(t)\right)v_k(u)v_\ell(w)\right|\right|_2\\
&\quad=\left|\left|\hat{c}_i\hat{v}_i-v_i\right|\right|_2+\left|\left|\hat{c}_j\hat{v}_j-v_j\right|\right|_2+\left|\left|\hat{c}_k\hat{v}_k-v_k\right|\right|_2+\left|\left|\hat{c}_\ell\hat{v}_\ell-v_\ell\right|\right|_2\\
&\quad=O_P\left(N^{-1/2}\right).
\end{aligned}$$

Hence (6.4) is proven which also completes the proof of Lemma 6.2. □

Lemma 6.3. *If Assumptions 1.1, 1.2, and 1.5 hold, then*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{F}}_n\hat{\mathbf{D}}_n^T=o_P\left(1\right).$$

Proof. We see from (6.1) that an element of $\mathbf{F}_n\mathbf{D}_n^T$ can be written in the form $\sqrt{\lambda_i\lambda_j\lambda_k}\xi_i^{(n)}\xi_j^{(n)}\xi_k^{(n)}$, where $\mathbf{F}_n = \left(\langle X_n^c, v_1\rangle, \quad \langle X_n^c, v_2\rangle, \quad \dots, \quad \langle X_n^c, v_p\rangle\right)^T$. We observe that $E\xi_i^{(n)}\xi_j^{(n)}\xi_k^{(n)} = 0$, so using the central limit theorem, we have

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{F}_n\mathbf{D}_n^T = O_P\left(N^{-1/2}\right).$$

Repeating the arguments in the proof (6.3), one can verify that

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\langle X_n-\bar{X}, \hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X}, \hat{c}_j\hat{v}_j\rangle\langle X_n-\bar{X}, \hat{c}_k\hat{v}_k\rangle - \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\langle X_n^c, v_k\rangle\Big) = o_P(1). \tag{6.6}$$

Since random signs do not affect convergence to zero, the proof is complete. □

Lemma 6.4. *If Assumptions 1.1, 1.2, and 1.5 hold, then*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{F}}_n\hat{\mathbf{F}}_n^T - \mathbf{\Lambda} = o_P(1),$$

where $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$.

Proof. By (6.1), an element of $\mathbf{F}_n\mathbf{F}_n^T$ is of the form $\sqrt{\lambda_i\lambda_j}\xi_i^{(n)}\xi_j^{(n)}$. Since $E\xi_i^{(n)}\xi_j^{(n)} = 1\{i=j\}$, according to the law of large numbers we have

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{F}_n\mathbf{F}_n^T - \mathbf{\Lambda} = o_P(1).$$

Thus it suffices to demonstrate that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n-\bar{X}, \hat{v}_i\rangle\langle X_n-\bar{X}, \hat{v}_j\rangle - \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\right) = o_P(1). \tag{6.7}$$

Since random signs do not affect convergence to zero, multiplying $\hat{v}_i$ by $\hat{c}_i$ and $\hat{v}_j$ by $\hat{c}_j$ will not affect convergence when $i \neq j$. If $i = j$, then $\hat{c}_i\hat{c}_j = \hat{c}_i^2 = 1$. Therefore, it suffices to show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n-\bar{X}, \hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X}, \hat{c}_j\hat{v}_j\rangle - \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\right) = o_P(1). \tag{6.8}$$

One can show (6.8) in exactly the same way we established (6.3) in the proof of Lemma 6.2. This completes the proof. □

Lemma 6.5. *If Assumptions 1.1, 1.2, and 1.5 hold, then*

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{F}}_n = o_P(1).$$

Proof. Using (6.1), an element of $\mathbf{F}_n$ has the form $\sqrt{\lambda_i}\xi_i^{(n)}$, so the law of large numbers implies that

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{F}_n = o_P(1).$$

The proof will be completed by establishing that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{F}_n - \hat{\mathbf{F}}_n\right) = o_P(1). \tag{6.9}$$

We express (6.9) componentwise and obtain

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n - \bar{X}, \hat{v}_i\rangle\right) = o_P(1). \tag{6.10}$$

Since random signs do not affect convergence to zero, it suffices to show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right) = o_P(1). \tag{6.11}$$

We will establish (6.11) in two steps. We will show that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n^c, \hat{c}_i\hat{v}_i\rangle\right) = o_P(1). \tag{6.12}$$

Then, we will establish that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, \hat{c}_i\hat{v}_i\rangle - \langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right) = o_P(1). \tag{6.13}$$

Using the central limit theorem in Hilbert spaces with Lemma 6.1 we conclude

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, v_i\rangle - \langle X_n^c, \hat{c}_i\hat{v}_i\rangle\right)\right| &\le \left|\left|\frac{1}{N}\sum_{n=1}^{N}X_n^c(t)\left(v_i - \hat{c}_i\hat{v}_i\right)\right|\right|_1 \\ &\le \left|\left|\frac{1}{N}\sum_{n=1}^{N}X_n^c(t)\right|\right|_2 ||v_i - \hat{c}_i\hat{v}_i||_2 \\ &= O_P\left(N^{-1}\right),\end{aligned}$$

and by the same arguments we have

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N}\left(\langle X_n^c, \hat{c}_i\hat{v}_i\rangle - \langle X_n - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right)\right| &= \left|\langle \mu_X - \bar{X}, \hat{c}_i\hat{v}_i\rangle\right| \\ &\le \left|\left|\left(\mu_X(t) - \bar{X}(t)\right)\hat{c}_i\hat{v}_i(t)\right|\right|_1 \\ &\le \left|\left|\mu_X(t) - \bar{X}(t)\right|\right|_2 \\ &= o_P(1).\end{aligned}$$

□

Lemma 6.6. *If Assumptions 1.1, 1.2, and 1.5 hold, then*

$$\hat{\mathbf{M}} - \mathbf{M} = o_P(1).$$

where $\hat{\mathbf{M}} = N^{-1} \sum_{n=1}^{N} \hat{\mathbf{D}}_n$ *and* $\mathbf{M} = E(\mathbf{D}_n)$.

Proof. An arbitrary element of $\hat{\mathbf{D}}_n$ is of the form

$$\frac{1}{N} \sum_{n=1}^{N} \langle X_n - \bar{X}, \hat{v}_i \rangle \langle X_n - \bar{X}, \hat{v}_j \rangle.$$

Since this is exactly the same as the form of an arbitrary element of $\hat{\mathbf{F}}_n \hat{\mathbf{F}}_n^T$, Lemma 6.6 follows from the proof of Lemma 6.4. Note in particular that when $i \neq j$, the sum converges to zero and is unaffected by signs, and when $i = j$, the signs cancel each other out. For this reason, $\boldsymbol{\zeta}\mathbf{M} = \mathbf{M}$, rendering it unnecessary to multiply $\mathbf{M}$ by $\boldsymbol{\zeta}$ in the statement of the lemma. □

Lemma 6.7. *If Assumptions 1.1, 1.2, and 1.5 hold, then*

$$\left(\left(\frac{\hat{\mathbf{Z}}^T \hat{\mathbf{Z}}}{N} \right) - \begin{pmatrix} \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} & \mathbf{0}_{r \times p} & \mathbf{M} \\ \mathbf{0}_{p \times r} & \boldsymbol{\Lambda} & \mathbf{0}_{p \times 1} \\ \mathbf{M}^T & \mathbf{0}_{1 \times p} & 1 \end{pmatrix} \right) = o_P(1).$$

Proof. This follows immediately from Lemmas 6.2–6.6. □

We will now use Lemma 6.7 to separate our estimate, $\hat{\mathbf{A}}$, of $\tilde{\mathbf{A}}$ from the estimates of the other parameters in (1.11).

Lemma 6.8. *If Assumptions 1.1–1.5 hold, then*

$$\boldsymbol{\zeta}\sqrt{N}\hat{\mathbf{A}} - N^{-1/2} \sum_{n=1}^{N} \varepsilon_n^{**} \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T \right)^{-1} \boldsymbol{\zeta} \left(\hat{\mathbf{D}}_n - \mathbf{M} \right) = o_P(1).$$

Proof. Let

$$\mathbf{C} = \begin{pmatrix} \boldsymbol{\zeta}\mathbf{G}\boldsymbol{\zeta} & \mathbf{0}_{r \times p} & \mathbf{M} \\ \mathbf{0}_{p \times r} & \boldsymbol{\Lambda} & \mathbf{0}_{p \times 1} \\ \mathbf{M}^T & \mathbf{0}_{1 \times p} & 1 \end{pmatrix}.$$

Using the fact that $\boldsymbol{\zeta}\mathbf{M} = \mathbf{M}$, one can verify via matrix multiplication that

$$\mathbf{C}^{-1} = \begin{pmatrix} \boldsymbol{\zeta} \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T \right)^{-1} \boldsymbol{\zeta} & \mathbf{0}_{r \times p} & -\boldsymbol{\zeta} \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T \right)^{-1} \boldsymbol{\zeta}\mathbf{M} \\ \mathbf{0}_{p \times r} & \boldsymbol{\Lambda}^{-1} & \mathbf{0}_{p \times 1} \\ -\mathbf{M}^T \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T \right)^{-1} & \mathbf{0}_{1 \times p} & 1 + \mathbf{M}^T \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T \right)^{-1} \mathbf{M} \end{pmatrix}.$$

Since $N^{-1/2}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$ is bounded in probability, by (4.1) and Lemma 6.7 we have

$$\sqrt{N}\begin{pmatrix}\hat{\mathbf{A}}\\ \hat{\mathbf{B}}-\tilde{\mathbf{B}}\\ \hat{\mu}-\mu\end{pmatrix}-\mathbf{C}^{-1}N^{-1/2}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}=o_P\left(1\right). \tag{6.14}$$

We observe that $\mathbf{C}^{-1}N^{-1/2}\hat{\mathbf{Z}}^T\boldsymbol{\varepsilon}^{**}$ can be expressed as

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\begin{pmatrix}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta} & \mathbf{0}_{r\times p} & -\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\\ \mathbf{0}_{p\times r} & \boldsymbol{\Lambda}^{-1} & \mathbf{0}_{p\times 1}\\ -\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1} & \mathbf{0}_{1\times p} & 1+\mathbf{M}^T\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\mathbf{M}\end{pmatrix}\begin{pmatrix}\hat{\mathbf{D}}_n\\ \hat{\mathbf{F}}_n\\ 1\end{pmatrix}. \tag{6.15}$$

Notice that the first $r=p(p+1)/2$ elements of the vector in (6.15) are given by

$$\begin{aligned}&N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\quad \mathbf{0}_{r\times p}\quad -\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right)\begin{pmatrix}\hat{\mathbf{D}}_n\\ \hat{\mathbf{F}}_n\\ 1\end{pmatrix}\\ &=N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\mathbf{M}\right)\\ &=N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right).\end{aligned}$$

Therefore

$$\sqrt{N}\hat{\mathbf{A}}-N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\boldsymbol{\zeta}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right)=o_P\left(1\right). \tag{6.16}$$

The result is now obtained by multiplying (6.16) on the left by $\boldsymbol{\zeta}$. □

Lemma 6.9. *If Assumptions 1.1–1.5 hold, then*

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\boldsymbol{\zeta}\left(\hat{\mathbf{D}}_n-\mathbf{M}\right)\xrightarrow{\mathcal{D}}N\left(0,\tau^2\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\right),$$

where

$$\tau^2=\sigma^2+\sum_{i=p+1}^{\infty}b_i^2\lambda_i$$

and $\sigma^2=\operatorname{var}\varepsilon_n$.

Proof. We prove this lemma in three steps. First we establish that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\left(\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\mathbf{M}\right)-\left(\mathbf{D}_n-\mathbf{M}\right)\right)=o_P\left(1\right). \tag{6.17}$$

In the second step we prove that

$$N^{-1/2}\sum_{n=1}^{N}(\mathbf{D}_n-\mathbf{M})\left(\varepsilon_n^{**}-\varepsilon_n^{*}-\sum_{i=1}^{p}b_i\langle\bar{X}-\mu_X,\hat{c}_i\hat{v}_i\rangle\right)=o_P(1) \tag{6.18}$$

and

$$N^{-1/2}\sum_{n=1}^{N}(\mathbf{D}_n-\mathbf{M})\langle\bar{X}-\mu_X,\hat{c}_i\hat{v}_i\rangle=o_P(1). \tag{6.19}$$

Combining (6.17), (6.18), and (6.19) we obtain immediately that

$$N^{-1/2}\sum_{n=1}^{N}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\left(\varepsilon_n^{**}\left(\boldsymbol{\zeta}\hat{\mathbf{D}}_n-\mathbf{M}\right)-\varepsilon_n^{*}(\mathbf{D}_n-\mathbf{M})\right)=o_P(1).$$

Therefore, the lemma will be established by the third step:

$$N^{-1/2}\sum_{n=1}^{N}\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\varepsilon_n^{*}(\mathbf{D}_n-\mathbf{M})\xrightarrow{\mathcal{D}}N\left(0,\tau^2\left(\mathbf{G}-\mathbf{M}\mathbf{M}^T\right)^{-1}\right). \tag{6.20}$$

We will now proceed to prove (6.17). The left side of (6.17) can be expressed elementwise as

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X},\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_i\rangle\langle X_n^c,v_j\rangle\right)=o_P(1), \tag{6.21}$$

so it is sufficient to show that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_i\rangle\langle X_n^c,v_j\rangle\right)=O_P\left(N^{-1/2}\right) \tag{6.22}$$

and

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\left(\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle X_n-\bar{X},\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\langle X_n^c,\hat{c}_j\hat{v}_j\rangle\right)=o_P(1). \tag{6.23}$$

The left side of (6.22) is

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)+N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,v_j\rangle\left(\langle X_n^c,\hat{c}_i\hat{v}_i\rangle-\langle X_n^c,v_i\rangle\right).$$

It follows from Assumptions 1.1–1.4 that both sets of random functions $\{\varepsilon_n X_n^c(t)X_n^c(s), 1\le n\le N\}$ and $\{X_n^c(u)X_n^c(t)X_n^c(s), 1\le n\le N\}$ are independent and identically distributed with zero mean so by the central limit theorem in Hilbert spaces we have

$$\left\|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n X_n^c(t)X_n^c(s)\right\|_2=O_P(1)\quad\text{and}\quad\left\|N^{-1/2}\sum_{n=1}^{N}X_n^c(u)X_n^c(t)X_n^c(s)\right\|_2=O_P(1). \tag{6.24}$$

Next we write that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)=\delta_1+\delta_2+\delta_3+\delta_4,$$

where, by (6.24), Lemma 6.1 and repeated applications of the Cauchy-Schwarz inequality, we have

$$\begin{aligned}|\delta_1|&=\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\left|\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_nX_n^c(t)X_n^c(s)\hat{c}_i\hat{v}_i(t)\left(\hat{c}_j\hat{v}_j(s)-v_j(s)\right)\right|\right|_1\\&\le\left|\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_nX_n^c(t)X_n^c(s)\right|\right|_2||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$\begin{aligned}|\delta_2|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\left|\left|N^{-1/2}\sum_{n=1}^{N}X_n^c(u)X_n^c(t)X_n^c(s)\right|\right|_2\left|\left|\sum_{k=p+1}^{\infty}b_kv_k(u)\right|\right|_2||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$\begin{aligned}|\delta_3|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=1}^{p}b_k\langle X_n^c,v_k-\hat{c}_k\hat{v}_k\rangle\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\sum_{k=1}^{p}|b_k|\left|\left|N^{-1/2}\sum_{n=1}^{N}X_n^c(t)X_n^c(s)X_n^c(w)\right|\right|_2||v_k(w)-\hat{c}_k\hat{v}_k(w)||_2\,||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1}\right),\end{aligned}$$

and

$$\begin{aligned}|\delta_4|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=1}^{p}b_k\langle\bar{X}-\mu_X,\hat{c}_k\hat{v}_k\rangle\langle X_n^c,\hat{c}_i\hat{v}_i\rangle\left(\langle X_n^c,\hat{c}_j\hat{v}_j\rangle-\langle X_n^c,v_j\rangle\right)\right|\\&\le\left|\sum_{k=1}^{p}b_k\langle\bar{X}-\mu_X,\hat{c}_k\hat{v}_k\rangle\right|\left|\left|N^{-1/2}\sum_{n=1}^{N}X_n^c(t)X_n^c(s)\right|\right|_2||\hat{c}_j\hat{v}_j(s)-v_j(s)||_2\\&=O_P\left(N^{-1/2}\right).\end{aligned}$$

Similarly,

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,v_j\rangle\left(\langle X_n^c,\hat{c}_i\hat{v}_i\rangle-\langle X_n^c,v_i\rangle\right)=o_P\left(1\right),$$

and therefore (6.22) is proven.

We now establish (6.23). The left side of (6.23) is equal to

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle+N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n^c,\hat{c}_j\hat{v}_j\rangle\langle\mu_X-\bar{X},\hat{c}_i\hat{v}_i\rangle.$$

We write that

$$N^{-1/2}\sum_{n=1}^{N}\varepsilon_n^{**}\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle=\delta_5+\delta_6+\delta_7+\delta_8,$$

where, by the central limit theorem in Hilbert spaces, Lemma 6.1, and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}|\delta_5|&=\left|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\\&\le\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left\|N^{-1/2}\sum_{n=1}^{N}\varepsilon_n\left(X_n(s)-\bar{X}(s)\right)\right\|_2\\&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$\begin{aligned}|\delta_6|&=\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\\
&\le\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|N^{-1/2}\sum_{n=1}^{N}\sum_{k=p+1}^{\infty}b_k\langle X_n^c,v_k\rangle\langle X_n-\bar{X},\hat{c}_i\hat{v}_i\rangle\right|\\
&=\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|N^{-1/2}\sum_{n=1}^{N}\int_0^1\int_0^1X_n^c(t)\left(X_n(s)-\bar{X}(s)\right)\hat{v}_i(s)\sum_{k=p+1}^{\infty}b_kv_k(t)\,ds\,dt\right|\\
&=\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|N^{-1/2}\sum_{n=1}^{N}\int_0^1\int_0^1\left(X_n(t)-\bar{X}(t)\right)\left(X_n(s)-\bar{X}(s)\right)\hat{v}_i(s)\sum_{k=p+1}^{\infty}b_kv_k(t)\,ds\,dt\right|\\
&=N^{1/2}\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|\int_0^1\int_0^1\hat{c}(t,s)\hat{v}_i(s)\sum_{k=p+1}^{\infty}b_kv_k(t)\,ds\,dt\right|\\
&=N^{1/2}\hat{\lambda}_i\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|\int_0^1\hat{v}_i(t)\sum_{k=p+1}^{\infty}b_kv_k(t)\,dt\right|\\
&=N^{1/2}\hat{\lambda}_i\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left|\int_0^1\sum_{k=p+1}^{\infty}b_kv_k(t)\left(\hat{v}_i(t)-\hat{c}_iv_i(t)\right)\,dt\right|\\
&\le N^{1/2}\hat{\lambda}_i\left|\langle\mu_X-\bar{X},\hat{c}_j\hat{v}_j\rangle\right|\left\|\sum_{k=p+1}^{\infty}b_kv_k(t)\right\|_2||\hat{v}_i(t)-\hat{c}_iv_i(t)||_2\\
&=O_P\left(N^{-1/2}\right),\end{aligned}$$

$$
\begin{aligned}
|\delta_7| &= \left| N^{-1/2} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \\
&\leq \left| \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \left\| N^{-1/2} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k X_n^c(t) \left(X_n(s) - \bar{X}(s) \right) \right\|_2 ||v_k(t) - \hat{c}_k \hat{v}_k(t)||_2 \\
&= O_P\left(N^{-1/2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
|\delta_8| &= \left| N^{-1/2} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \langle X_n - \bar{X}, \hat{c}_i \hat{v}_i \rangle \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \\
&\leq \left| \langle \mu_X - \bar{X}, \hat{c}_j \hat{v}_j \rangle \right| \left| \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \right| \left\| N^{-1/2} \sum_{n=1}^{N} \left(X_n(s) - \bar{X}(s) \right) \right\|_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}
$$

This proves (6.23), which also completes the proof of (6.21) and hence (6.17).

We proceed to the second step, which is the proof of (6.18) and (6.19). We express (6.18) elementwise as

$$
N^{-1/2} \sum_{n=1}^{N} \left(\langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle - \lambda_i 1\{i = j\} \right) \left(\sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \right) = o_P(1). \tag{6.25}
$$

We observe that by the central limit theorem in Hilbert spaces and Lemma 6.1 we have

$$
\begin{aligned}
\left| N^{-1/2} \sum_{n=1}^{N} \left(\sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \right) \right| &\leq \left\| N^{-1/2} \sum_{n=1}^{N} X_n^c(t) \right\|_2 \sum_{k=1}^{p} |b_k| \, ||v_k(t) - \hat{c}_k \hat{v}_k(t)||_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\left| N^{-1/2} \sum_{n=1}^{N} \langle X_n^c, v_i \rangle \langle X_n^c, v_j \rangle \left(\sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \right) \right| \\
&\qquad \leq \sum_{k=1}^{p} |b_k| \left\| N^{-1/2} \sum_{n=1}^{N} X_n^c(t) X_n^c(s) X_n^c(w) \right\|_2 ||v_k(w) - \hat{c}_k \hat{v}_k(w)||_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}
$$

This proves (6.25) and hence (6.18). Next, we establish (6.19). We can express (6.19) elementwise as

$$
N^{-1/2} \sum_{n=1}^{N} \left(\langle X_n^c, v_k \rangle \langle X_n^c, v_\ell \rangle - \lambda_k 1\{k = \ell\} \right) \langle \bar{X} - \mu_X, \hat{c}_i \hat{v}_i \rangle = o_P(1). \tag{6.26}
$$

Using the previous arguments, one can easily verify (6.26), establishing (6.19).

We will now finish the proof of the lemma by establishing (6.20) as the third step. Using Assumptions 1.1, 1.3, and (1.4), we see that ε_n^* has mean zero and variance given by

$$
\begin{aligned}
E\left(\varepsilon_n^*\right)^2 &= E\left(\varepsilon_1^2\right) + E\left(\sum_{i=p+1}^{\infty}\sum_{j=p+1}^{\infty} b_i b_j \langle X_n^c, v_i\rangle\langle X_n^c, v_j\rangle\right) \\
&= \sigma^2 + \sum_{i=p+1}^{\infty} b_i^2 E\left(\langle X_n^c, v_i\rangle^2\right) \\
&= \sigma^2 + \sum_{i=p+1}^{\infty} b_i^2 \lambda_i. \\
&= \tau^2
\end{aligned}
$$

Therefore, $\varepsilon_n^*\left(\mathbf{D}_n - \mathbf{M}\right)$ is an iid sequence with mean zero and variance $\tau^2\left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)$. The central limit theorem now proves (6.20), completing the proof of the lemma. □

Lemma 6.10. *If Assumptions 1.2–1.5 are satisfied, then*

$$
\begin{pmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{B}} \\ \hat{\mu} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \tilde{\mathbf{B}} \\ \mu \end{pmatrix} = O_P\left(N^{-1/2}\right). \tag{6.27}
$$

In particular, we have

$$
\|b_k v_k(t) - \hat{b}_k \hat{v}_k(t)\|_2 = O_P\left(N^{-1/2}\right) \tag{6.28}
$$

and

$$
\|\hat{a}_{i,j}\hat{v}_i(t)\hat{v}_j(s)\|_2 = O_P\left(N^{-1/2}\right), \tag{6.29}
$$

where $\hat{a}_{i,j}$ and $\hat{b}_i$ are defined by

$$
\hat{\mathbf{A}} = \operatorname{vech}\left(\left\{\hat{a}_{i,j}\left(2 - 1\{i=j\}\right), 1 \le i \le j \le p\right\}^T\right) \quad \textit{and} \quad \hat{\mathbf{B}} = \begin{pmatrix}\hat{b}_1, & \hat{b}_2, & \ldots, & \hat{b}_p\end{pmatrix}^T.
$$

Proof. Lemmas 6.8 and 6.9 imply that $\hat{\mathbf{A}} = O_P\left(N^{-1/2}\right)$. According to (6.14) and (6.15) we can prove that

$$
\hat{\mathbf{B}} - \tilde{\mathbf{B}} = O_P\left(N^{-1/2}\right), \tag{6.30}
$$

by showing that

$$
\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^{**}\mathbf{\Lambda}^{-1}\hat{\mathbf{F}}_n = O_P\left(N^{-1/2}\right)
$$

or equivalently that

$$
\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^{**}\langle X_n - \bar{X}, \hat{v}_i\rangle = O_P\left(N^{-1/2}\right). \tag{6.31}
$$

We note that

$$
\frac{1}{N}\sum_{n=1}^{N} \varepsilon_n^{**}\langle X_n - \bar{X}, \hat{v}_i\rangle = \delta_9 + \delta_{10} + \delta_{11} + \delta_{12},
$$

where, following the arguments in the proof of Lemma 6.9, one can verify that

$$|\delta_9| = \left| \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n \langle X_n - \bar{X}, \hat{v}_i \rangle \right| O_P\left(N^{-1/2}\right),$$

$$|\delta_{10}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=p+1}^{\infty} b_k \langle X_n^c, v_k \rangle \langle X_n - \bar{X}, \hat{v}_i \rangle \right| = O_P\left(N^{-1/2}\right),$$

$$|\delta_{11}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k \rangle \langle X_n - \bar{X}, \hat{v}_i \rangle \right| = O_P\left(N^{-1/2}\right),$$

and

$$|\delta_{12}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \langle X_n - \bar{X}, \hat{v}_i \rangle \right| = O_P\left(N^{-1/2}\right).$$

This proves (6.31) and hence (6.30).

To complete the justification of (6.27), we need to show that

$$\hat{\mu} - \mu = O_P\left(N^{-1/2}\right). \tag{6.32}$$

Due to (6.14) and (6.15), (6.32) will be established by proving that

$$\frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^{**} \left(-\mathbf{M}^T \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1} \hat{\mathbf{D}}_n + 1 + \mathbf{M}^T \left(\mathbf{G} - \mathbf{M}\mathbf{M}^T\right)^{-1} \mathbf{M} \right) = O_P\left(N^{-1/2}\right). \tag{6.33}$$

To prove (6.33), it is sufficient to show

$$\frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^{**} \hat{\mathbf{D}}_n = O_P\left(N^{-1/2}\right) \tag{6.34}$$

and

$$\frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^{**} = O_P\left(N^{-1/2}\right). \tag{6.35}$$

Due to Lemma 6.9, (6.35) implies (6.34), so we prove only (6.35). We write that

$$\frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^{**} = \delta_{13} + \delta_{14} + \delta_{15} + \delta_{16},$$

where, by the central limit theorem in Hilbert spaces and Lemma 6.1, we have

$$|\delta_{13}| = \left| \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n \right| = O_P\left(N^{-1/2}\right),$$

$$|\delta_{14}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=p+1}^{\infty} b_k \langle X_n^c, v_k \rangle \right| \leq \left\| \frac{1}{N} \sum_{n=1}^{N} X_n^c(t) \right\|_2 \left\| \sum_{k=p+1}^{\infty} b_k v_k(t) \right\|_2 = O_P\left(N^{-1/2}\right),$$

$$|\delta_{15}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle X_n^c, v_k - \hat{c}_k \hat{v}_k(t) \rangle \right| = O_P\left(N^{-1}\right),$$

and

$$|\delta_{16}| = \left| \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{p} b_k \langle \bar{X} - \mu_X, \hat{c}_k \hat{v}_k \rangle \right| = O_P\left(N^{-1/2}\right).$$

This proves (6.35), which establishes (6.32) and completes the proof of (6.27).

Using (6.27) and Lemma 6.1, we will now show (6.28) and (6.29). We conclude from (6.27) that

$$\hat{b}_i - \hat{c}_i b_i = O_P\left(N^{-1/2}\right) \quad \text{and} \quad \hat{a}_{i,j} = O_P\left(N^{-1/2}\right).$$

Now, Lemma 6.1 yields that

$$\begin{aligned} \|b_k v_k(t) - \hat{b}_k \hat{v}_k(t)\|_2 &\le \|b_k(v_k(t) - \hat{c}_k \hat{v}_k(t))\|_2 + \|(b_k \hat{c}_k - \hat{b}_k)\hat{v}_k(t)\|_2 \\ &\le |b_k| \|v_k(t) - \hat{c}_k \hat{v}_k(t)\|_2 + |b_k \hat{c}_k - \hat{b}_k| \\ &= O_P\left(N^{-1/2}\right). \end{aligned}$$

Similarly,

$$\|\hat{a}_{i,j} \hat{v}_i(t) \hat{v}_j(s)\|_2 = O_P\left(N^{-1/2}\right).$$

This proves (6.28) and (6.29) and completes the proof of the lemma. $\square$

Lemma 6.11. *If Assumptions 1.1–1.5 are satisfied, then*

$$\hat{\tau}^2 - \tau^2 = O_P\left(N^{-1/2}\right).$$

Proof. Since

$$\frac{1}{N} \sum_{n=1}^{N} \varepsilon_n^{*2} - \tau^2 \xrightarrow{a.s.} 0,$$

it is enough to show that

$$\frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n^2 - \varepsilon_n^{*2}\right) = O_P\left(N^{-1/2}\right). \tag{6.36}$$

Since

$$\frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n^2 - \varepsilon_n^{*2}\right) = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right)\left(\hat{\varepsilon}_n + \varepsilon_n^*\right) = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \hat{\varepsilon}_n + \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \varepsilon_n^*,$$

(6.36) follows from

$$\left| \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \varepsilon_n^* \right| = O_P\left(N^{-1/2}\right) \tag{6.37}$$

and

$$\left| \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\varepsilon}_n - \varepsilon_n^*\right) \hat{\varepsilon}_n \right| = O_P\left(N^{-1/2}\right). \tag{6.38}$$

We decompose (6.37) as

$$\frac{1}{N}\sum_{n=1}^{N}\left(\hat{\varepsilon}_n-\varepsilon_n^*\right)\varepsilon_n^*=\eta_1+\eta_2+\eta_3,$$

where

$$\eta_1=\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^*\left(\mu-\hat{\mu}\right),$$

$$\eta_2=\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^*\sum_{i=1}^{p}\Big(b_i\langle X_n^c,v_i\rangle-\hat{b}_i\langle X_n-\bar{X},\hat{v}_i\rangle\Big),$$

$$\eta_3=\frac{1}{N}\sum_{n=1}^{N}\varepsilon_n^*\sum_{i=1}^{p}\sum_{j=i}^{p}(2-1\{i=j\})\left(a_{i,j}\langle X_n^c,v_i\rangle\langle X_n^c,v_j\rangle-\hat{a}_{i,j}\langle X_n-\bar{X},\hat{v}_i\rangle\langle X_n-\bar{X},\hat{v}_j\rangle\right).$$

It is clear that $\eta_1=O_P\left(N^{-1}\right)$. We also see that $\eta_2=\eta_{2,1}+\eta_{2,2}+\eta_{2,3}+\eta_{2,4}$, where

$$\eta_{2,1}=\frac{1}{N}\sum_{n=1}^{N}Y_n\sum_{i=1}^{p}\Big(b_i\langle X_n^c,v_i\rangle-\hat{b}_i\langle X_n-\bar{X},\hat{v}_i\rangle\Big),$$

$$\eta_{2,2}=-\frac{1}{N}\sum_{n=1}^{N}\mu\sum_{i=1}^{p}\Big(b_i\langle X_n^c,v_i\rangle-\hat{b}_i\langle X_n-\bar{X},\hat{v}_i\rangle\Big),$$

$$\eta_{2,3}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{\ell=1}^{p}b_\ell\langle X_n^c,v_\ell\rangle\sum_{i=1}^{p}\Big(b_i\langle X_n^c,v_i\rangle-\hat{b}_i\langle X_n-\bar{X},\hat{v}_i\rangle\Big),$$

$$\eta_{2,4}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{\ell=1}^{p}\sum_{k=\ell}^{p}(2-1\{k=\ell\})a_{\ell,k}\langle X_n^c,v_\ell\rangle\langle X_n^c,v_k\rangle\sum_{i=1}^{p}\Big(b_i\langle X_n^c,v_i\rangle-\hat{b}_i\langle X_n-\bar{X},\hat{v}_i\rangle\Big).$$

Applying (6.28) and the central limit theorem in Hilbert spaces we obtain that

$$\begin{aligned}
|\eta_{2,1}| &= \left| \frac{1}{N} \sum_{n=1}^{N} Y_n \sum_{i=1}^{p} \Big(b_i \langle X_n^c, v_i \rangle - \hat{b}_i \langle X_n - \bar{X}, \hat{v}_i \rangle \Big) \right| \\
&\leq \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n \Big(b_i X_n^c(t) v_i(t) - \hat{b}_i \left(X_n(t) - \bar{X}(t) \right) \hat{v}_i(t) \Big) \right\|_1 \\
&\leq \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n X_n(t) \Big(b_i v_i(t) - \hat{b}_i \hat{v}_i(t) \Big) \right\|_1 \\
&\quad + \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n \Big(b_i \mu_X(t) v_i(t) - \hat{b}_i \bar{X}(t) \hat{v}_i(t) \Big) \right\|_1 \\
&\leq \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n X_n(t) \right\|_2 \left\| b_i v_i(t) - \hat{b}_i \hat{v}_i(t) \right\|_2 \\
&\quad + \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n \bar{X}(t) \Big(\hat{b}_i \hat{v}_i(t) - b_i v_i(t) \Big) \right\|_1 \\
&\quad + \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n b_i v_i(t) \left(\bar{X}(t) - \mu_X(t) \right) \right\|_1 \\
&\leq \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n X_n(t) \right\|_2 \left\| b_i v_i(t) - \hat{b}_i \hat{v}_i(t) \right\|_2 \\
&\quad + \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n \bar{X}(t) \right\|_2 \left\| \hat{b}_i \hat{v}_i(t) - b_i v_i(t) \right\|_2 \\
&\quad + \sum_{i=1}^{p} \left\| \frac{1}{N} \sum_{n=1}^{N} Y_n b_i v_i(t) \right\|_2 \left\| \bar{X}(t) - \mu(t) \right\|_2 \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}$$

In a like manner, one can verify that $\eta_{2,i} = O_P\left(N^{-1/2}\right), i = 2, 3, 4$. This proves that $\eta_2 = O_P\left(N^{-1/2}\right)$. In a similar fashion, one can show that $\eta_3 = O_P\left(N^{-1/2}\right)$. This proves (6.37). Following the previous arguments, one can establish (6.38), completing the proof of the lemma. □

References

Borggaard, C. and H. Thodberg (1992). Optimal minimal neural interpretation of spectra. *Analytical Chemistry 64*, 545–551.

Bosq, D. (2000). *Linear Processes in Function Spaces*. New York: Springer.

Cai, T. and P. Hall (2006). Prediction in functional linear regression. *The Annals of Statistics 34*, 2159–2179.

Cardot, H., F. Ferraty, A. Mas, and P. Sarda (2003). Testing hypothesis in the functional linear model. *Scandinavian Journal of Statistics 30*, 241–255.

Cardot, H. and P. Sarda (2011). Functional linear regression. In F. Ferraty and Y. Romain (Eds.), *The Oxford Handbook of Functional Data Analysis*, pp. 21–46. Oxford University Press.

Dauxois, J., A. Pousse, and Y. Romain (1982). Asymptotic theory for principal component analysis of a vector random function. *Journal of Multivariate Analysis 12*, 136–154.

Fitzhugh, H. A. (1976). Analysis of growth curves and strategies for altering their shapes. *Journal of Animal Science 33*, 1036–1051.

Frank, I. E. and J. H. Friedman (1993). A statistical view of some chemometrics regression tools. *Technometrics 35*, 109–135.

Hall, P. and J. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*, 70–91.

Kirkpatrick, M. and N. Heckman (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of Mathematical Biology 27*, 429–450.

Mas, A. and B. Pumo (2011). Linear processes for functional data. In F. Ferraty and Y. Romain (Eds.), *The Oxford Handbook of Functional Data Analysis*, pp. 47–71. Oxford University Press.

Müller, H.-G. and Y. Zhang (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics 61*, 1064–1075.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Springer.

Shen, Q. and J. Faraway (2004). An f test for linear models with functional responses. *Statistica Sinica 14*, 1239–1257.

Wold, S. (1993). Discussion: Pls in chemical practice. *Technometrics 35*, 136–139.

Yao, F. and H.-G. Müller (2010). Functional quadratic regression. *Biometrika 97*, 49–64.